Predictive Maintenance of Mining Centrifuges Using Machine and Deep Learning Models

Doaa Ahmad Alqaraleh [a,1], Sami Salama Hussen Hajjaj[b,c,2], Hassan Mohamed[b,d,3], Mohd Radzi Aridi[d,4] Mohd Zafri bin Baharuddin[e,f,5]

[a]College of Graduate Studies (COGS), Universiti Tenaga Nasional (UNITEN), 43000 Kajang, Selangor, Malaysia

[b]School of Computing and Artificial Intelligence, Faculty of Engineering and Technology, Sunway University, No. 5, Jalan Universiti, Bandar Sunway, 47500 Selangor Darul Ehsan, Malaysia.

[c]Research Centre For Human-Machine Collaboration (HUMAC), Faculty of Engineering and Technology, Sunway University, No. 5, Jalan Universiti, Bandar Sunway, 47500 Selangor Darul Ehsan, Malaysia.

[d]Institute of Sustainable Energy (ISE), Universiti Tenaga Nasional (UNITEN), 43000 Kajang, Selangor, Malaysia

[e]Mechanical Engineering Department, Universiti Tenaga Nasional (UNITEN), 43000 Kajang, Selangor, Malaysia

[f]Institute of Energy Infrastructure (IEI) College of Engineering Universiti Tenaga Nasional,43000 Kajang, Selangor, Malaysia

Emails:

Doua.ahmad90@yahoo.com , samihajjaj@sunway.edu.my , mhassan@uniten.edu.my,

radzi@uniten.edu.my , Zafr@uniten.edu.my

Abstract

This study aims to facilitate the achievement of Sustainable Development Goals (SDG 9: Industry, Innovation, and Infrastructure & SDG 13: Climate Action) using intelligent maintenance solutions. Attaining operational and environmental sustainability in the mining sector poses significant challenges that necessitate innovative solutions to enhance equipment efficiency and minimize unforeseen failures. This research formulates a predictive maintenance system utilizing artificial intelligence and machine learning techniques, grounded in the analysis of sensor data from industrial centrifuges. The models evaluated were Deep Neural Networks (DNN), Long Short-Term Memory (LSTM), K-Nearest Neighbors (KNN), Naïve Bayes (NB), Logistic Regression (LR), and XGBoost. XGBoost demonstrated exceptional proficiency in managing time-series data, achieving the greatest accuracy of 99.80%, followed by LSTM. These results underscore the potential of machine learning techniques to optimize predictive maintenance, reduce energy and resource consumption, enhance operational efficiency, and promote sustainable industrial practices.

Keywords

Predictive Maintenance, Machine Learning, Deep Learning, Industrial Centrifuge, Mining Industry, Sustainable Development Goals, and industrial innovation.

1. Introduction

Particularly in the mining industry, constant production and cost control in the modern corporate environment depend on efficient machinery maintenance and running. Industrial centrifuges are crucial tools for mineral processing and separation since they use centrifugal force to separate valuable components from ore mixtures. The dependability of these machines determines direct effects on production rates and resource recovery efficiency. But

wear and tear can lead to unexpected breakdowns that result in major financial losses and more repair expenses [1].

Unexpected centrifuge failures not only affect repair costs but also cause manufacturing disturbances, affecting operational objectives. Predictive maintenance (PdM) has evolved as a solution using advanced technology to detect and correct system faults before they start [2]. Unlike traditional maintenance methods, which are either reactive or based on set schedules, PdM alters how mining operations preserve their critical equipment. Predictive maintenance systems use machine learning (ML) and deep learning (DL) techniques to identify minor anomalies that may indicate an impending failure, utilizing vast sensor data [3]. These computational methods shine in spotting complex correlations within multivariate data streams, enabling early identification of possible errors sometimes weeks or months before a failure occurs.

Working under challenging conditions, the mining industry accelerates equipment degradation unlike in other sectors. Smart predictive maintenance systems will significantly benefit this industry [4]. By using sensor networks, edge computing, and advanced analytics, the integration makes it easier to see how the centrifuges work and gives important information for making maintenance schedules more efficient and reducing downtime [1]. Accurate failure prediction and maintenance planning optimization enable mining companies to lower costs and increase general operational efficiency [5]. Moreover, predictive maintenance lowers material waste and prolongs equipment lifetime, supporting sustainability.

Predictive maintenance is among the best approaches to maximize industrial operations and lower downtime. Reactive use of predictive maintenance techniques causes a major paradigm change in industrial operations management. Underlining the need for automated systems to manage vast amounts of industrial process data, [6] underlines the part predictive maintenance plays in reducing costs and preserving efficiency. Extreme working conditions and high mining machinery equipment costs lead to particularly difficult maintenance. For instance, [7] applied sequence mining technology to assess mining truck conditions, enhancing maintenance planning and demonstrating the need for particular solutions in the mining sector. Predictive maintenance is essential since dangerous working conditions and expensive manufacturing losses resulting from broken equipment call for it , In recent findings, vibration analysis was emphasized as the most significant tool of rotating machinery's predictive maintenance, considering its capability in early fault diagnosis [8].
In response, Khalil and Rostam introduced a semi-automated system based on Fast Fourier Transform (FFT) for extracting frequency-domain characteristics of vibration signals, which are trained and classified based on Support Vector Machines (SVM) and ensemble schemes, and secured accuracy of more than 90% [9].
These results illustrate the value of combining machine learning and signal processing so that efficient and reliable predictive maintenance can be achieved.

The application of ML techniques in predictive maintenance has transformed the mining industry. [10]looked at conventional maintenance techniques, which often lead to inefficiencies and high costs; they concluded predictive maintenance is a better option. Underlining the need for first-rate data, [11] looked at IoT and artificial intelligence

applications in PdM. Similarly, [12]looked at the rail sector to show how ML models might increase system dependability, directly influencing mining activities.

For machine learning models, including Random Forest, XGBoost, and neural networks, predictive maintenance applications have shown success. [13]presented Random Forest algorithms for early failure detection in high-dimensional datasets for which they were able to XGBoost has also improved predictive accuracy by reducing false failure projections; this is an essential capacity in mining environments with high data complexity. Other studies, like [14], stress the need for machine learning in high-risk areas. They talk about maintenance strategies that lower financial losses and raise operational safety.

Deep learning has especially evolved predictive maintenance capabilities even more for complex equipment like centrifuges. Long short-term memory (LSTM) networks have shown successful analysis of time-series data from industrial sensors, drastically lowering running costs and raising equipment dependability[15, 16].CNNs are especially good at analyzing sensor data, especially for early degradation in centrifuge components depending on vibration patterns [17]. Despite these advances, predictive maintenance in mining operations still presents challenges using ML and DL approaches. Data quality and availability present important difficulties since imbalanced, noisy, or incomplete data can influence model performance [18]. Moreover, the "black box" of deep learning models characterizes interpretability, which complicates maintenance teams' trust and activities depending on predictions [19]. Researchers [20] have underlined the need for automated, dependable inspection systems fit for the challenging conditions of the mining sector.

This paper presents a new predictive maintenance framework especially designed for industrial centrifuges used in the mining industry. Our system differs from earlier work depending on synthetic or lab-based data in utilizing real sensor data extracted from an ABB 800xa DCS. We further combine several ML and DL models, e.g., DNN, LSTM, KNN, NB, LR, and XGBoost, and combine these in a voting-based ensemble, aiming at increasing the accuracy of predictions. Such an approach, alongside pragmatic considerations of utilizing the solution, covers data reliability issues, interpretability of the model, and operational scalability, and provides a realistic and adaptive solution for addressing maintenance issues in rugged industrial conditions.
One area of development is predictive maintenance models for industrial centrifuges used in mining. Artificial intelligence-driven predictive maintenance solutions help mining operations increase operational efficiency and equipment dependability. This work investigates ML and DL models, including DNN, LSTM, k-NN, Naive Bayes, Logistic Regression, and XGBoost, to build a strong predictive maintenance system. The proposed system collects real sensor data from centrifuges, applies ML/DL algorithms for health monitoring, and utilizes these results through a voting system to enhance prediction accuracy. Using real-world sensor data from an ABB 800xa DCS, this work has significant relevance to practical use, unlike earlier studies dependent on synthetic data.

More and more, the mining industry is using machine learning and deep learning techniques to enable industrial centrifuges to be predictive and maintenance-friendly. Advanced deep learning architectures like convolutional neural networks (CNNs) and long short-term memory networks (LSTMs) help mining companies predict when equipment will break down before it happens. Using real-time operational parameters and past sensor data, these models maximize maintenance schedules, reduce expected downtime, and enhance decision-making processes.

Notwithstanding their promise, some challenges still exist, including data quality issues, model interpretability, computational resource constraints, and seamless integration into current industry systems. Data source heterogeneity, changes in sensor precision, and noise interference further complicate the building of robust predictive models. Moreover, the adoption of AI-driven maintenance plans demands significant infrastructure investment and workforce upskilling to ensure efficient application.

As the mining sector follows its digital transformation route, predictive maintenance is probably going to become a pillar of operational efficiency, workplace safety, and environmental sustainability. This work aims to contribute to this shifting terrain by designing, developing, and thoroughly evaluating a complete predictive maintenance system suited for industrial centrifugation processes, The following Figure 1 illustrates an image of the industrial centrifuge that will be used in our study. The work especially employs a diverse suite of machine and deep learning models—including deep neural networks (DNN), LSTMs, k-nearest neighbours (KNN), naïve Bayes (NB), logistic regression (LR), and XGBoost—here discussed—to increase the accuracy and dependability of failure predictions.

Figure 1: Image of the industrial centrifuge used in this study.

With real-world sensor data collection, the proposed system combines feature engineering techniques to increase predictive accuracy. By methodically comparing model performance under many running conditions, this work provides empirical insights into the most effective strategies for industrial centrifuge maintenance. Furthermore, the study helps to close the present knowledge gap and so supports better equipment lifetime and cost-efficiencies in the mining industry by offering an adaptive, scalable, and interpretable predictive maintenance framework.

2. Methodology:

This inquiry is based on the multi-stage unified cross-industry mining process (CRISP-DM) technique. Determining the issues with industrial maintenance and investigating possible solutions is the first step in the business understanding stage. The following step is data comprehension, which focuses on acquiring and understanding the data collection. The second step in the process is processing the data to prepare it for analysis. It then proceeds to the modelling phase, where techniques of machine learning and deep learning are applied. The evaluation follows to assess their effectiveness, followed by the implementation stage, which uses the top model to predict maintenance in real-time; Figure 2.1 shows the framework of our research

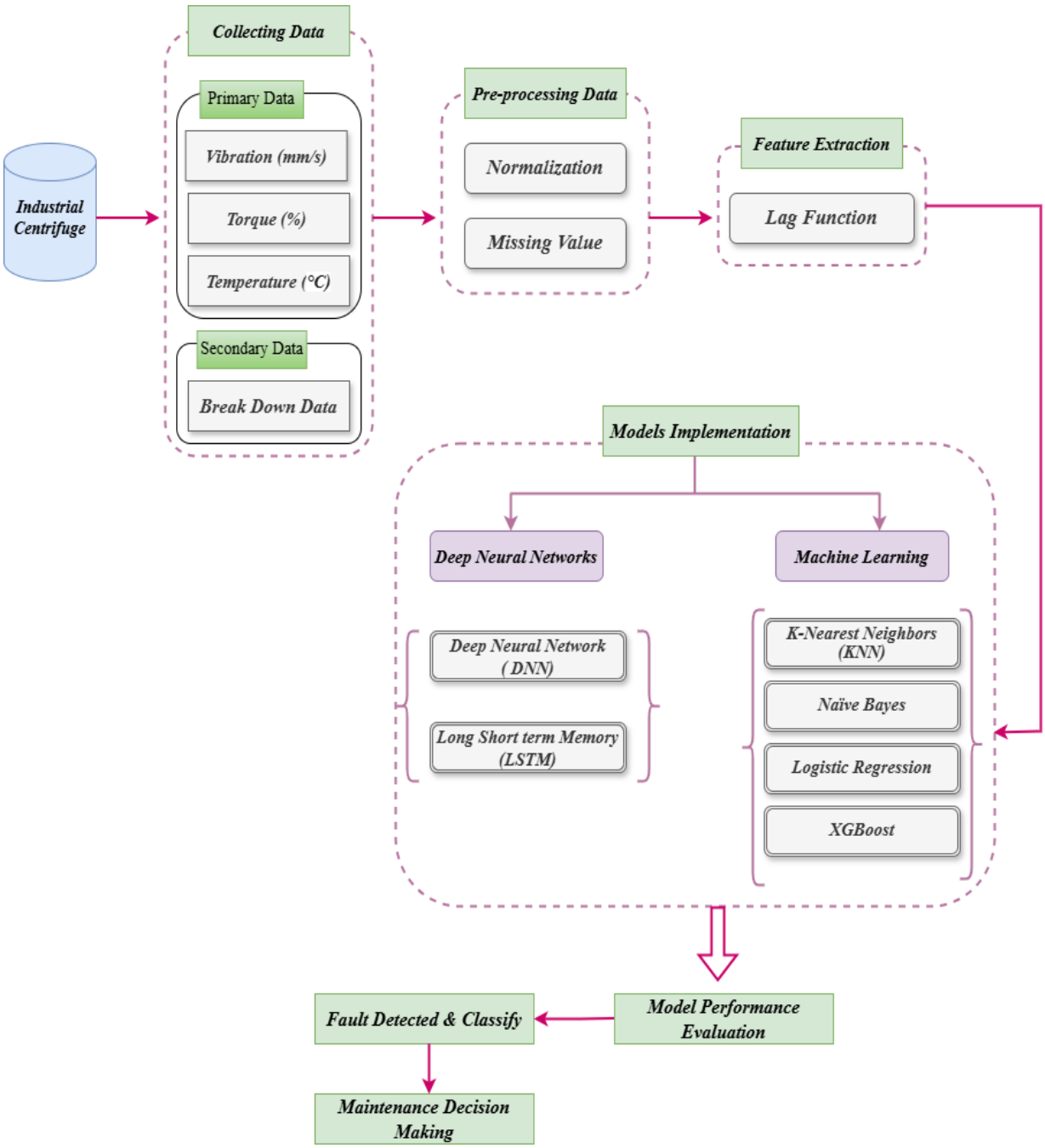


Figure 2.1: The architecture of I-PdM platform

2.1 Collecting data:

An ABB 800xa series distributed control system (DCS) manages all the data from different types of field sensors, such as analogue, digital, or contact-based ones. To improve facility performance, this system logs data, displays metrics on tailored graphical interfaces, and archives them for subsequent analysis. Careful distribution of the sensors guarantees accurate monitoring. Heat sensors are incorporated in the centrifugal bearings to prevent too

much overheating, while lubricant sensors track tank conditions to ensure they stay within safe working limits. The vibration sensor searches for unusual mechanical disturbances while the torque sensor monitors the gear output shaft to ensure the centrifuge runs within its limits. Using infrared technology, the overheating sensor in the hydraulic clutch monitors precisely, free from human contact. This complete method guarantees maintenance operations are improved, and the proactive identification of future problems is made possible by ensuring the acquisition and analysis of previous data. Figure 2.2 presents the industrial centrifuge along with the sensor mounted on it. Figure 2.3 illustrates a graph depicting the readings from various sensors recorded over a 12-month observation period. Additionally, Table 1 provides a sample of the larger dataset, showcasing the key attributes on which this work is based.

Table 1: sample of Sensors readings

Date	*Torque*	*Vibration*	*Main bearing Temp - Solid end*	*Main bearing Temp - Liquid End*
9/5/2022 1:00	59.98	0.29	58.08	57.09
9/5/2022 1:01	59.99	0.35	57.90	57.08
9/5/2022 1:02	59.81	0.21	57.82	57.12
9/5/2022 1:03	59.63	0.23	57.84	57.12
9/5/2022 1:04	60.14	0.24	57.91	57.13
9/5/2022 1:05	60.20	0.22	57.95	57.13
9/5/2022 1:06	59.83	0.29	57.90	57.14
9/5/2022 1:07	59.74	0.21	57.94	57.14
9/5/2022 1:08	59.86	0.29	57.80	57.15
9/5/2022 1:09	59.95	0.37	57.70	57.16
9/5/2022 1:10	59.84	0.35	57.69	57.14
9/5/2022 1:11	60.00	0.29	57.63	57.13
9/5/2022 1:12	60.11	0.30	57.63	57.15
9/5/2022 1:13	59.59	0.35	57.67	57.14
9/5/2022 1:14	60.42	0.29	57.64	57.11
9/5/2022 1:15	60.12	0.37	57.79	57.08

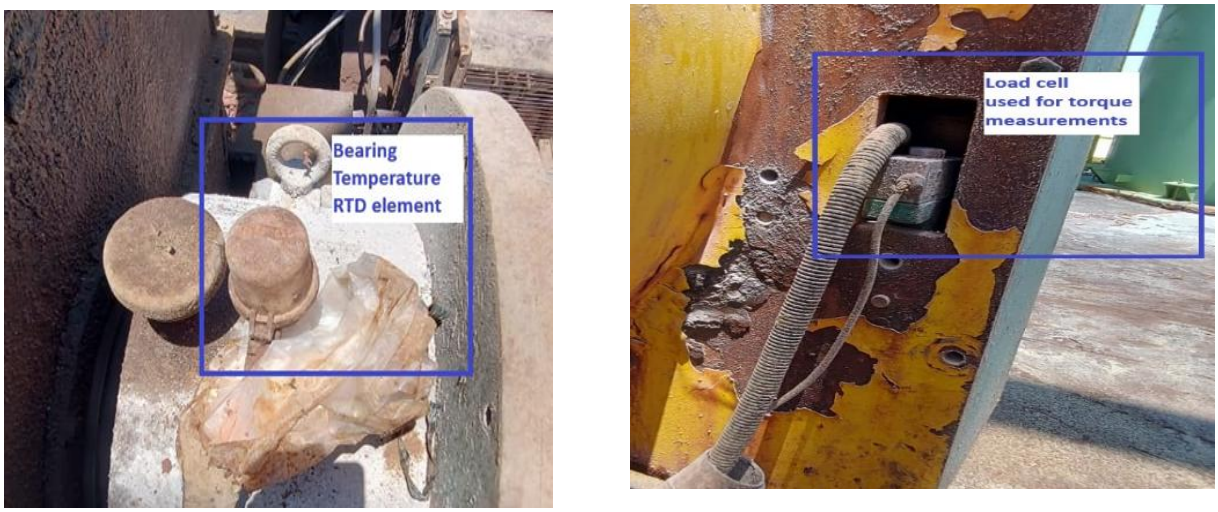
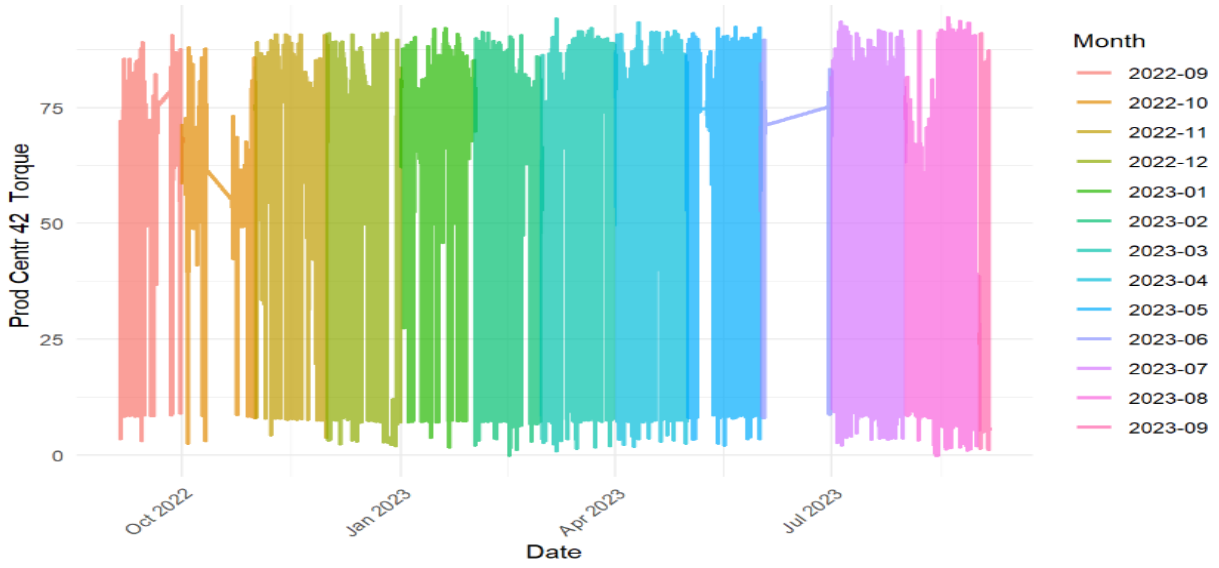
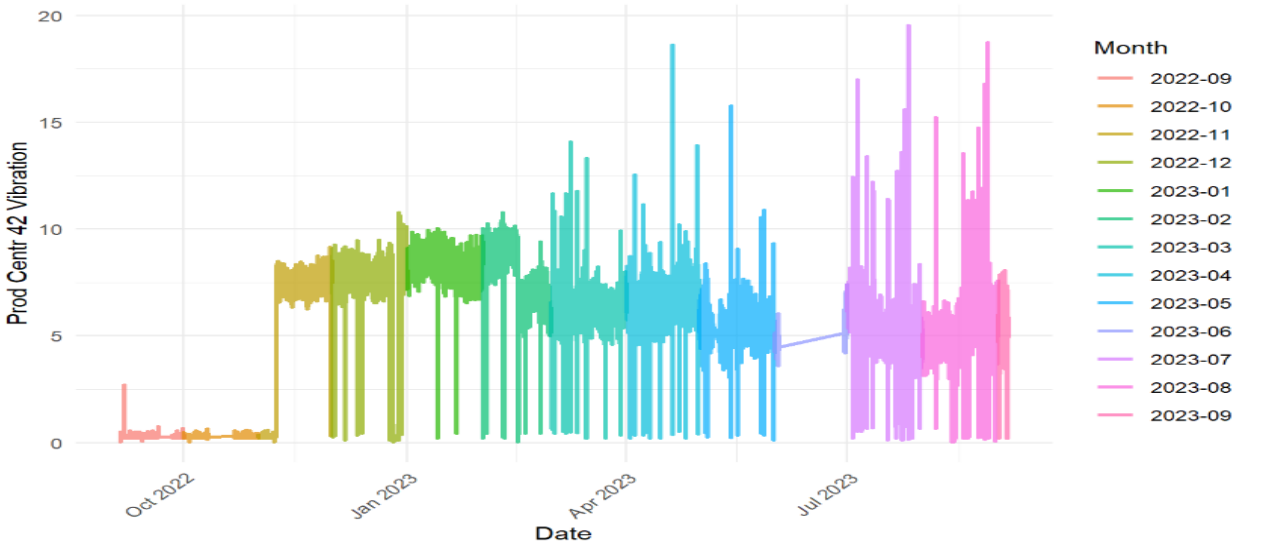


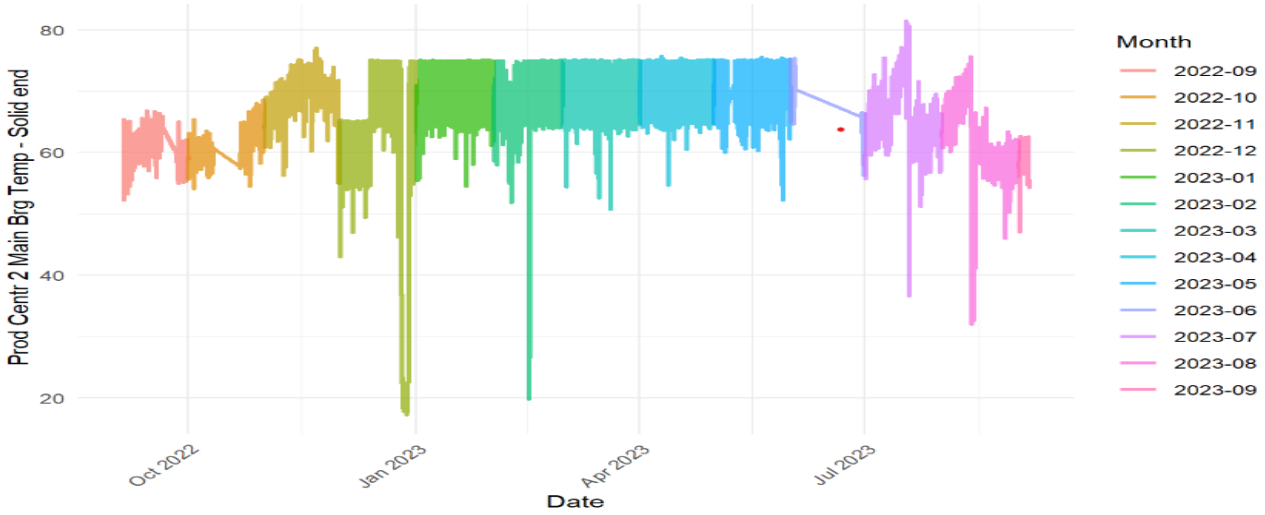
Figure 2.2: Some of the Sensors located at centrifuges


Prod Centr 42 Torque
75
50
25
0
Oct 2022
Jan 2023
Apr 2023
Jul 2023
Date
Month
2022-09
2022-10
2022-11
2022-12
2023-01
2023-02
2023-03
2023-04
2023-05
2023-06
2023-07
2023-08
2023-09

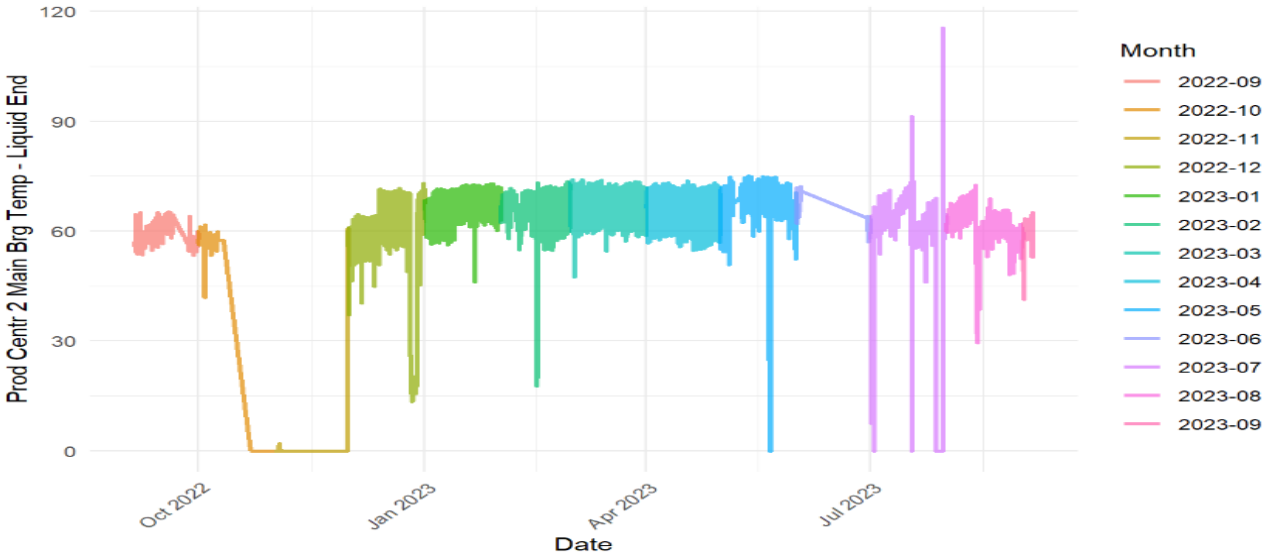
(1)


Prod Centr 42 Vibration
20
15
10
5
0
Oct 2022
Jan 2023
Apr 2023
Jul 2023
Date
Month
2022-09
2022-10
2022-11
2022-12
2023-01
2023-02
2023-03
2023-04
2023-05
2023-06
2023-07
2023-08
2023-09

(2)


Prod Centr 2 Main Brg Temp - Solid end
80
60
40
20
Oct 2022
Jan 2023
Apr 2023
Jul 2023
Date
Month
2022-09
2022-10
2022-11
2022-12
2023-01
2023-02
2023-03
2023-04
2023-05
2023-06
2023-07
2023-08
2023-09

(3)



(4)

Figure 2.3: Sensor readings for an industrial centrifuge over 12 months

Successful data collection relies on identifying centrifuge issues. Overheating, vibration unbalance, belt wear, oil flow disruptions, and hydraulic coupling difficulties are mechanical causes of downtime. Problems with motors, sensors, and other electrical and mechanical devices can stop operations and need to be looked into right away to avoid harm and keep data accurate.

Sensor alarms monitor performance: vibration alarm at 12 mm/s, torque alarm at 85% power, and oil temperature alarm at 85°C signal critical problems. In the interests of productivity as well as assisting research goals, our DCS system has remote-adjustment and real-time monitor sensors.

2.2 Data Preparation:

2.2.1 Data Cleaning

Data preparation began with a thorough dataset examination to find and fix any missing information. Data integrity and consistency were crucial before proceeding with further processing. The computer looked over the dataset and then indicated how many values from every column were lacking. The first analysis revealed the spectrum of missing data, therefore enabling a structured process to remedy shortfalls.

Except for the "Label" field, the column mean was utilized to impute missing values for numerical columns. We used this imputation because it had minimal impact on the dataset's statistical characteristics or distribution. Keeping the dataset whole and ready for analysis helps reduce the bias resulting from missing data. The program uses placeholders to provide context-sensitive control over non-numeric columns.

Following default disablement, inclusion allows text-based or category data manipulation. The modular character of the technique allows more complex imputation algorithms to be included as necessary.

Following this phase, the program reviewed and reported the revised count of missing values for each column, thereby verifying the missing data management procedure. This verification stage ensured that the program had filled in any missing data before processing.

The dataset underwent annotations for categorization following data cleansing. The function label data labels the data entries for operating parameters such as torque, power, temperature, and vibration. This search is designed to identify instances of parameter

overruns. The row labelled "Fault," meaning a system issue when a parameter went above its threshold.

But if every measurement came out normal, the row was labelled "Normal." This labelling process came from calling the apply method on every DataFrame entry to activate label_data. The orderly filling of the "Label" column prepared the dataset for model training and analysis. This method for automatic classification helped organize the dataset so that prediction models could find problems and make smart choices while the system was being inspected and fixed.

Table 2: Binary classification of the centrifuge operating condition for the ML and DL models.

Description	Range	Critical Limit	Model Classification
Vibration sensor	0-20 mm/sec	12mm/sec	0: Normal operating condition <12mm/sec 1: Critical operating condition ≥ 12mm/sec
Torque sensors	0%-100% power%	85 power%	0: Normal operating condition <85 power% 1: Critical operating condition ≥ 85 power%
Bearing temperature 1	0-100 degrees C	85 degrees C	0: Normal operating condition <85 degrees C 1: Critical operating condition ≥ 85 degrees C
Bearing temperature 2	0-100 degrees C	0-100 degrees C	0: Normal operating condition <85 degrees C 1: Critical operating condition ≥ 85 degrees C

2.2.2 engineering features:

Future values in time series are strongly influenced by past values. Lagging manufacturing metrics like torque, power, vibration, and temperature allowed one to track serial relationships. Delaying variable starting values by a specified number of periods (lags) allowed new characteristics derived from past data.

This work applied a 10-step lag window. We conducted this work to identify short- and medium-term trends in the data. With a latency of 10, the model detects trends or anomalies in system behaviours using patterns and connections across several data points. Production parameter changes require time to affect system performance; hence, the industrial process under analysis had ten delays.

A 10-lag gives the data set qualities that enable the model to learn from past adjustments. This is needed for predictive modelling—more especially, sequential data sets and machine learning techniques, including RNNs and gradient boosting models. Lagged variables contextualize observations to assist in finding anomalies and forecasting.

The script adds one for each feature lagged one-time step to ten steps back, ten columns for each measure. Knowing this helps the model to spot trends and produce more accurate forecasts.

2. Datetime Conversion, Component Extraction

Pandas assured consistency by converting the 'Date' column into a datetime object, therefore simplifying temporal data handling, Upon conversion, more datetime object columns were utilized to mark the year, month, day, hour, minute, and day of the week. Their inclusion provides various advantages.

- Data segmentation shows seasonal trends and operating cycles over daily, weekly, or monthly periods.

• Found trends: Some events or anomalies show increased frequency on specific days or times. Through merging temporal variables, the method detects concealed trends and patterns. Enhancement of Model Interpretation: Certain machine learning techniques value time-related characteristics above timestamps. The time series study made possible by this dataset structure helps measure production patterns.

Training and Evaluating Models:

2.3.1Model Selection

A lot of different models are trained and tested during the model selection step. These models include K-Nearest Neighbors (KNN), Naive Bayes (NB), Logistic Regression (LR), XGBoost, Deep Neural Network (DNN), and Long Short-Term Memory (LSTM). The selection of these models was meticulous, considering their various approaches and varying levels of efficiency when applied to time series datasets.

The chosen algorithms, such as K-Nearest Neighbors (KNN), Naive Bayes, Logistic Regression, XGBoost, Deep Neural Networks (DNN), and Long Short-Term Memory (LSTM) networks, are some of the best at classifying things. Each model has its own set of computational features that help with the many-sided problems of predictive maintenance. These include handling decision boundaries that aren't linear and figuring out how complex temporal dependencies The table1 provided in the appendix offers an overview of each algorithm used along with its corresponding mathematical model.

Because it can handle non-linear data and put complicated patterns into groups, the K-nearest neighbours (KNN) algorithm was chosen [21]. NB has shown that it can quickly sort data with a lot of dimensions, which is a task that comes up a lot in predictive maintenance applications[22]. The selection of LR was based on its ability to generate precise probabilities for categorization, which makes interpreting the results easier [23].

XGBoost stands out for its exceptional performance in several data science competitions and its capability to handle skewed data, a common occurrence in predictive maintenance settings[24]. When DNN was chosen, it was because it could automatically extract features and do very well in difficult classification tests [25]. Finally, the Long Short-Term Memory (LSTM) model was picked because it can effectively capture long-term relationships in sequential data. This makes it a good choice for analyzing time-series sensor data [26].

The selection of models was made to encompass a broad spectrum of methodologies, ranging from basic algorithms to sophisticated deep-learning techniques. This method allows a thorough evaluation of model performance using different approaches, which guarantees the discovery of the best model for the specific goal of predictive maintenance in industrial centrifuge operations [27].

2.3.2 training process

The dataset was first split into training and test sets to make sure that the model was trained correctly and that its performance on new data could be accurately judged. During the training phase, we implemented the k-fold cross-validation technique, specifically 5-fold cross-validation, to improve the robustness of the model and reduce overfitting. This method entailed the division of the training set into five distinct subsets. Each model was trained five times, with a different subset serving as the validation set in each iteration. This method aids in the selection of the most robust model and provides a more dependable estimate of the model's generalization performance [28].

The k-fold cross-validation technique is widely recognized in the field of machine learning for its capacity to conduct a comprehensive assessment of model performance, particularly in the presence of restricted data. It is essential in predictive maintenance

applications, where data quality and quantity can vary substantially, as it reduces bias and variance in model selection [29].

2.4 Evaluation Metrics

We used several metrics to evaluate the performance of the trained models: confusion matrix, accuracy, precision, recall, and F1 score [29]. The confusion matrix provides an overview of how the classification is done by showing the number of true positives, true negatives, false positives, and false negatives [30].

Accuracy is a metric that evaluates the overall correctness of forecasts, whereas precision, also known as positive predictive value, specifically measures the accuracy of positive predictions[31]. Recall, sometimes referred to as sensitivity, is the ability of the model to accurately identify all positive events [32]. The F1 score, calculated as the harmonic mean of precision and recall, offers a well-balanced evaluation of the model's performance [33]. Collectively, these metrics provide a thorough assessment of the effectiveness of the classification models. The metrics are precisely determined through the utilization of the following mathematical equations:

$$\boldsymbol{Accuracy} = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \times 100 \quad (1)$$

$$\boldsymbol{Precision} = \frac{True\ Positive}{True\ Positive + False\ Negative} \times 100 \quad (2)$$

$$\boldsymbol{Recall} = \frac{True\ Negative}{True\ Negative + False\ Positive} \times 100 \quad (3)$$

$$\boldsymbol{F1 - Score} = \frac{2 \times (Precision \times Recall)}{Precision \times Recall} \times 100 \quad (4)$$

3. Results

Our study evaluated seeral machine learning and deep learning models for the predictive maintenance of industrial centrifuges in the mining industry. We tested individual models and then proposed an ensemble method to enhance prediction accuracy.

Part 1: Individual Model Performance

We evaluated the following individual models: Deep Neural Networks (DNN), Long Short-Term Memory (LSTM), K-Nearest Neighbors (KNN), Naive Bayes (NB), Logistic Regression (LR), and XGBoost. Table 4 summarizes their performance metrics.

Table 4: Performance Metrics for Individual Models

Model	Accuracy	Precision (Normal/Fault)	Recall (Normal/Fault)	F1-Score (Normal/Fault)
DNN	95.73%	0.96 / 0.89	0.98 / 0.80	0.97 / 0.84
LSTM	95.86%	0.97 / 0.87	0.98 / 0.82	0.98 / 0.84
KNN	96.74%	0.98 / 0.89	0.98 / 0.89	0.98 / 0.89
NB	76.90%	0.98 / 0.39	0.74 / 0.91	0.84 / 0.54
LR	98.10%	0.99 / 0.95	0.99 / 0.93	0.99 / 0.94
XGBoost	99.00%	0.998 / 0.99	1.00 / 1.00	0.998 / 0.99

Figure 4.1 shows the accuracy rates of every model, so enabling a visual comparison of the accuracy of individual models. The XGBoost model displayed the best performance at 99.80%, while the Naïve Bayes (NB) model displayed the lowest performance at 76.90.

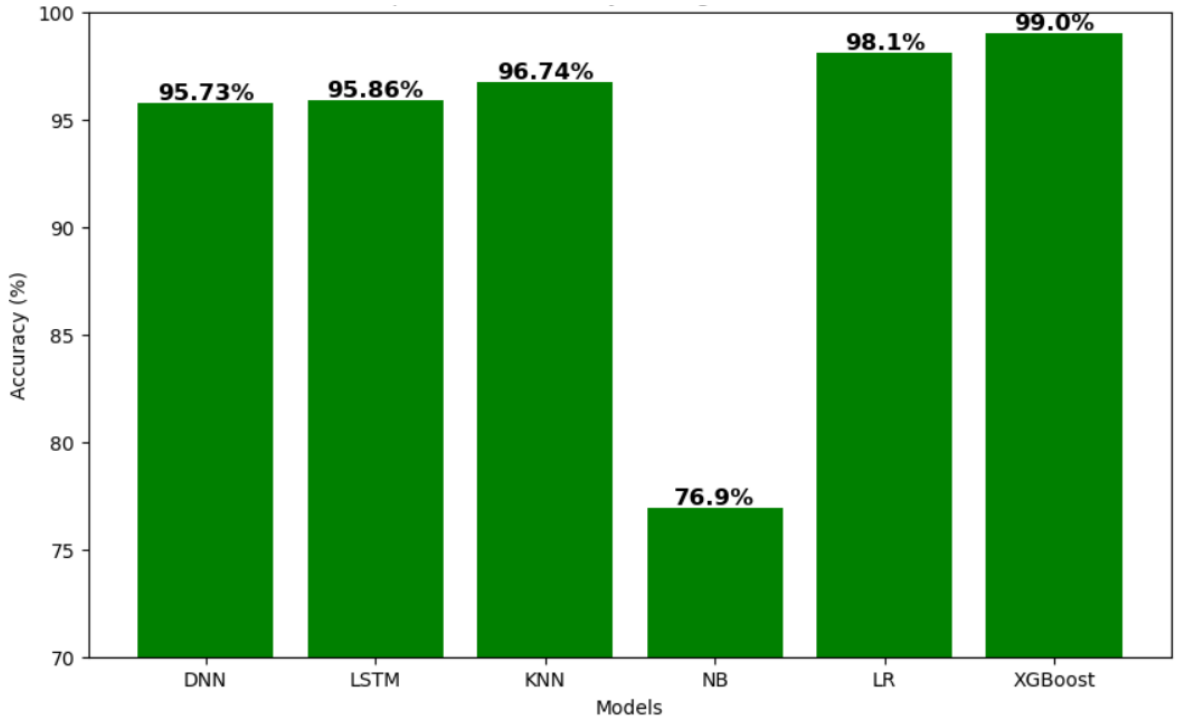


Figure 4.1: Accuracy Comparison of Individual Models

The deep learning models (LSTM and DNN) worked quite well, with LSTM performing slightly better than DNN because it can model temporal dependencies in sensor data. KNN also performed quite well, showing its strength in detecting patterns in multidimensional sensor data. On the other hand, the Naive Bayes model showed the worst performance, a behaviour that can be attributed to its assumption of feature independence. This assumption is unlikely to be held in the case of industrial sensor data, where complex and correlated relationships among variables often prevail. Logistic Regression achieved a high accuracy rate, keeping a good balance between precision and recall. Finally, XGBoost outperformed all the other models by a large margin, highlighting its capability to deal with high-dimensional and complicated datasets efficiently.

4. Discussion

Our results reveal the appreciable increases in predictive maintenance accuracy and dependability achieved using the integration of deep learning and machine learning models. Long Short-Term Memory (LSTM), K-Nearest Neighbors (KNN), naïve Bayes, logistic regression, and XGBoost—individual models—showed their dependability and robustness in evaluation. Every one of these models exposed special strengths and weaknesses that affected their generalizing and predictive accuracy. Based on the analysis in terms of predictive power and robustness, XGBoost and LSTM exceeded the other models, so obtaining good evaluation ratings.

As shown in **Figure 1**, each model has distinct strengths and weaknesses that influence its predictive accuracy and generalization capabilities. **XGBoost** is a good substitute for predictive maintenance fault detection since it shows amazing accuracy and recall and is well-known for managing structured data. Using its capacity to identify temporal dependencies, **LSTM** generated accurate equipment failure projections and performed remarkably well in time-series data pattern recognition.

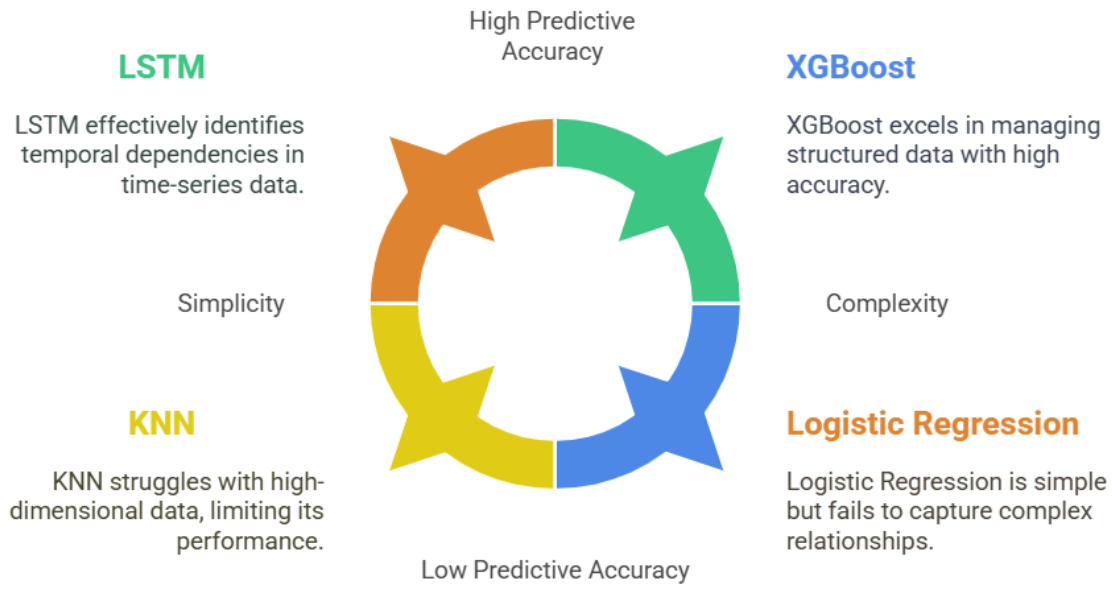

High Predictive Accuracy
LSTM
LSTM effectively identifies temporal dependencies in time-series data.
XGBoost
XGBoost excels in managing structured data with high accuracy.
Simplicity
Complexity
KNN
KNN struggles with high-dimensional data, limiting its performance.
Logistic Regression
Logistic Regression is simple but fails to capture complex relationships.
Low Predictive Accuracy

Figure 1: Comparative Analysis of Predictive Maintenance Models

By contrast, conventional machine learning models, including logistic regression, naïve Bayes, and KNN, show varied degrees of success. Though its simplicity in use, **KNN** suffered with high-dimensional data and produced less than ideal performance in demanding predictive maintenance environments. Because of their strong independence assumption, **naïve Bayes** show limits; hence, their classification accuracy is influenced even by their simplicity and computational efficiency. While **logistic regression** was understandable and efficient, it fell short of sufficiently reflecting complex relationships in the data, reducing the predictive accuracy in comparison to more sophisticated models.

The very high degree of accuracy and recall attained by the best-performing models helps maintenance teams to actively plan interventions and so enables early identification of operational anomalies. Our results also showed the ideal decision threshold at 0.0371, so balancing the trade-off between false positive and negative rates. False positives and false negatives decrease in small misclassification rates according to confusion matrices. This ability helps to avoid unplanned equipment breakdowns and optimize maintenance plans, helping to prevent unnecessary interventions and decrease running downtime.

Comparative Analysis and Theoretical Outcomes

Although many previous studies show the advantages of predictive maintenance, most have concentrated on single-model approaches. These models have certain benefits in terms of computational efficiency and interpretability even if they sometimes ignore entirely the possibilities of many algorithms. Our analysis reveals the corresponding strengths and weaknesses of particular models, guiding their applicability for predictive maintenance.

From a theoretical standpoint, our results add empirical data confirming the efficacy of advanced machine learning models, increasing the already increasing corpus of predictive maintenance research. The outcomes support the theory that deep learning and tree-based models—such as LSTM and XGBoost—have higher predictive capacity than more traditional

statistical models. These observations fit predictive maintenance theory, which emphasizes the need for using data-driven models to improve fault detection and lower running risks.

Real Results and Practical Uses in Industry

Apart from only theoretical advantages, this study has great pragmatic value for industrial maintenance programs. Modern industrial operations depend on predictive maintenance absolutely since equipment failures could cause significant financial losses and operational disturbance. Using exact, early fault detection systems, the adoption of high-performance predictive models offers the potential to improve maintenance policies.

In industrial environments, basic threshold-based alarms or regular inspections usually help to control maintenance schedules. Though rather effective, these conventional approaches lack the flexibility needed to control demanding and dynamic operational conditions. Using advanced predictive maintenance systems grounded on high-performance models such as XGBoost and LSTM, companies will be guided toward a more data-driven and proactive maintenance paradigm. Excellent accuracy and recall rates of these models show the ability of maintenance teams to better allocate resources, reducing needless inspections and interventions and concurrently minimizing catastrophic equipment failures.

Industries with high capital expenditures in heavy machinery, including mining, manufacturing, and energy generation, should especially find our results relevant. Predictive maintenance models will help these sectors increase operational efficiency, reduce unplanned downtime, and extend the lifetime of important machinery. Reducing resource waste and optimizing energy use enhances predictive maintenance capacity, supporting goals for sustainability.

Restraints and Future Directions of Research

This study has limits even if the results show great possibilities. Arab Potash Company's industrial centrifuges running under provided the whole basis for both training and evaluation. The results might thus be contextually limited and not always relevant for all sorts of industrial environments or equipment. Furthermore, gathered over twelve months, the dataset might not cover all possible operational scenarios or failure trends.

Future studies should try to overcome these constraints by including a more varied range of industrial equipment and extending the period of data collection to guarantee model robustness over several operational environments. Moreover, the integration of creative deep learning techniques such as reinforcement learning and transfer learning will help to improve things even if some models show exceptional predictive capacity. These techniques can improve model adaptability and learning efficiency, enabling predictive maintenance systems to run efficiently in many industrial environments needing little retraining. Future research should investigate hybrid methods combining deep learning architectures with conventional machine learning models to even more improve predictive performance.

Conclusion

Broadly speaking, This work emphasizes how well deep learning and machine learning models combined increase the predictive maintenance accuracy for industrial centrifuges. Restating their benefits in industry applications, individual model evaluation showed that XGBoost and LSTM advanced techniques provide better recall, precision, and general accuracy than conventional models. These results imply that, especially in industrial environments where equipment dependability is critical, acceptance of high-performance predictive models can greatly enhance the efficacy of predictive maintenance strategies. We suggest, as a future path of investigation, the aggregating of the strengths of several models, so enhancing predictive accuracy using a voting majority approach. Future studies can

investigate how aggregating several predictive models inside a single framework improves classification performance, transcending the constraints of conventional single-model techniques. Providing a reasonable and scalable solution for industrial operations, this creative approach supports better-informed decision-making, lowers maintenance costs, and raises operational effectiveness. Moreover, by tackling the difficulties of predictive maintenance in demanding industrial environments, this work provides important new perspectives on the body of knowledge already in use and provides a strong foundation for the next developments in smart maintenance technologies.

References:

[1] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications," *Mechanical systems and signal processing,* vol. 42, no. 1-2, pp. 314-334, 2014.

[2] R. Mobley, "An introduction to predictive maintenance Butterworth," ed: Heinemann, America, 2002.

[3] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical systems and signal processing,* vol. 20, no. 7, pp. 1483-1510, 2006.

[4] T. Zonta, C. A. Da Costa, R. da Rosa Righi, M. J. de Lima, E. S. Da Trindade, and G. P. Li, "Predictive maintenance in the Industry 4.0: A systematic literature review," *Computers & Industrial Engineering,* vol. 150, p. 106889, 2020.

[5] A. Mosallam, K. Medjaher, and N. Zerhouni, "Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction," *Journal of Intelligent Manufacturing,* vol. 27, pp. 1037-1048, 2016.

[6] P. Bastos, I. d. S. Lopes, and L. Pires, "A maintenance prediction system using data mining techniques," 2012.

[7] A. Kahraman, "Maintainability analysis of mining trucks with data analytics," 2018.

[8] W. Prommachan, P. Surin, P. Srinoi, and M. Pipathattakul, "Selection Criteria for Evaluating Predictive Maintenance Techniques for Rotating Machinery using the Analytic Hierarchical Process (AHP)," *Engineering, Technology & Applied Science Research,* vol. 14, no. 1, pp. 13058-13065, 2024.

[9] A. F. Khalil and S. Rostam, "Machine learning-based predictive maintenance for fault detection in rotating machinery: A case study," *Engineering, Technology & Applied Science Research,* vol. 14, no. 2, pp. 13181-13189, 2024.

[10] T. Zhu, Y. Ran, X. Zhou, and Y. Wen, "A survey of predictive maintenance: Systems, purposes and approaches," *arXiv preprint arXiv:1912.07383,* 2019.

[11] G. G. Samatas, S. S. Moumgiakmas, and G. A. Papakostas, "Predictive maintenance-bridging artificial intelligence and iot," in *2021 IEEE World AI IoT Congress (AIIoT)*, 2021: IEEE, pp. 0413-0419.

[12] N. Davari, B. Veloso, G. d. A. Costa, P. M. Pereira, R. P. Ribeiro, and J. Gama, "A survey on data-driven predictive maintenance for the railway industry," *Sensors,* vol. 21, no. 17, p. 5739, 2021.

[13] U. Farooq, M. Ademola, and A. Shaalan, "Comparative analysis of machine learning models for predictive maintenance of ball bearing systems," *Electronics,* vol. 13, no. 2, p. 438, 2024.

[14] S. Paraschos, I. Mollas, N. Bassiliades, and G. Tsoumakas, "Visiored: A visualisation tool for interpretable predictive maintenance," *arXiv preprint arXiv:2103.17003,* 2021.

[15] D. K. Sharma, S. Brahmachari, K. Singhal, and D. Gupta, "Data driven predictive maintenance applications for industrial systems with temporal convolutional networks," *Computers & Industrial Engineering,* vol. 169, p. 108213, 2022.

[16] S. Maheshwari, S. Tiwari, S. Rai, and S. V. D. P. Singh, "Comprehensive Study Of Predictive Maintenance In Industries Using Classification Models And LSTM Model," *arXiv preprint arXiv:2403.10259,* 2024.

[17] H. Wang, W. Zhang, D. Yang, and Y. Xiang, "Deep-learning-enabled predictive maintenance in industrial internet of things: methods, applications, and challenges," *IEEE Systems Journal,* vol. 17, no. 2, pp. 2602-2615, 2022.

[18] P. Nunes, J. Santos, and E. Rocha, "Challenges in predictive maintenance–A review," *CIRP Journal of Manufacturing Science and Technology,* vol. 40, pp. 53-67, 2023.

[19] M. Attaran and B. G. Celik, "Digital Twin: Benefits, use cases, challenges, and opportunities," *Decision Analytics Journal,* vol. 6, p. 100165, 2023.
[20] M. Pincheira, M. Antonini, and M. Vecchio, "Integrating the IoT and blockchain technology for the next generation of mining inspection systems," *Sensors,* vol. 22, no. 3, p. 899, 2022.
[21] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN classification with different numbers of nearest neighbors," *IEEE transactions on neural networks and learning systems,* vol. 29, no. 5, pp. 1774-1785, 2017.
[22] H. Zhang and D. Li, "Naïve Bayes text classifier," in *2007 IEEE international conference on granular computing (GRC 2007)*, 2007: IEEE, pp. 708-708.
[23] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013.
[24] C. T. G. C. XGBoost, "A scalable tree boosting system," in *Proc 22nd ACM SIGKDD Int Conf Knowl Discov Data Min*, 2016, pp. 785-794.
[25] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature,* vol. 521, no. 7553, pp. 436-444, 2015.
[26] N. Computation, "Long short-term memory," *Neural Comput,* vol. 9, pp. 1735-1780, 2016.
[27] T. P. Carvalho, F. A. Soares, R. Vita, R. d. P. Francisco, J. P. Basto, and S. G. Alcalá, "A systematic literature review of machine learning methods applied to predictive maintenance," *Computers & Industrial Engineering,* vol. 137, p. 106024, 2019.
[28] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," 2010.
[29] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information processing & management,* vol. 45, no. 4, pp. 427-437, 2009.
[30] T. Fawcett, "An introduction to ROC analysis," *Pattern recognition letters,* vol. 27, no. 8, pp. 861-874, 2006.
[31] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *arXiv preprint arXiv:2010.16061,* 2020.
[32] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PloS one,* vol. 10, no. 3, p. e0118432, 2015.
[33] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC genomics,* vol. 21, pp. 1-13, 2020.

Table 1: details each model and its corresponding mathematical equations

Model	Description	Mathematical Equation	Equation number
K-Nearest Neighbors (KNN)	Using the majority class (for classification) or the average value of the neighbours (for regression), K-Nearest Neighbors determines the "k" closest data points to the query point and generates a forecast. It works well for unknown distribution tiny datasets.	$(x) = \frac{1}{k} \sum_{i \in neighbors} Y_i$	(1)
Naive Bayes	Built on Bayes' Theorem, naive Bayes is a probabilistic classification method. It supposes, given the class label, characteristics to be conditionally independent. For text classification, it works well and is computationally efficient.	$CapCapP(\, y\|x) = \frac{P(y \prod_{i=1}^{n} P(x_i l\, y\,))}{P(x)}$	(2)
Logistic Regression	Binary categorization uses logistic regression. It uses a logistic function to translate the linear combination of input data into a probability, therefore modelling the likelihood of a particular class or occurrence.	$CapCapP(\, y = 1\|x\,) = \frac{1}{1 + \exp\big(-(w^T x + b)\big)1}$	(3)
XGBoost	Designed as an enhanced gradient-boosting method, XGBoost creates an ensemble of decision trees in which every tree fixes the mistakes of the others. High performance, scalability, and adaptability abound from it.	$f\,(x) = \sum_{k=1}^{k} \alpha_k h_k\,(x)$	(4)
Deep Neural Network (DNN)	Comprising several layers between the input and output layers, DNNs have data-capable complexity in capture ability. For jobs including picture and speech recognition, they are strong; but, they demand lots of data and computational capability.	$f(x) = \sigma(w_L \ldots \sigma(w_2\, \sigma\,(w_1 x + b_1) + b_2\,) \ldots) + b_L$	(5)
Long Short-Term Memory (LSTM)	LSTMs are a class of RNNs intended to represent time series and sequence data. For jobs involving sequential data, they are efficient since they keep a cell state capable of storing information over extended times.	$c_{\mathrm{t}} = f_{\mathrm{t}} \odot\ c_{\mathrm{t}-1} + i_{\mathrm{t}} \odot\ \tilde{\mathrm{c}}$	(6)